Neural networks - energy and robustness

Petra Vidnerová

Department of Artificial Intelligence, ÚI AV ČR

Jizerka Seminar, September, 2023

Outline

- very brief introduction to neural networks
- objectives of architecture selection - accuracy, energy, robustness
- robustness to noise and adversarial examples
- robustness of networks with approximate convolutional layers
- architecture selection - robustness evaluations pitfalls
- conclusion

Introduction

Neural networks

- hot topic nowadays
- image processing, signal analysis, large language models
- classification, regression, generative tasks

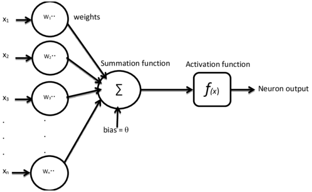
Our focus

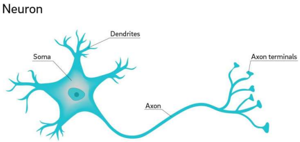
- image classification
- convolutional neural networks

Brief introduction to neural networks

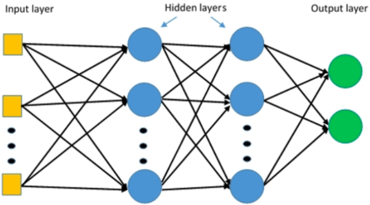
Artificial Neuron

- basic building block of all neural networks



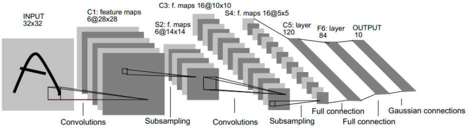


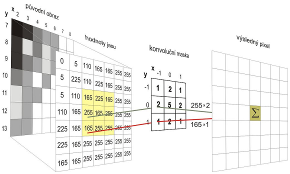
MultiLayer Perceptron



Brief introduction to neural networks

Convolutional Neural Networks

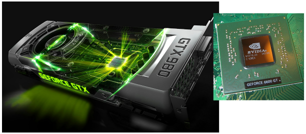




- 1994 LeNet5 (Yann LeCun)
- convolutional layers for feature extraction
- sub-sampling layers (max-pool layers)
- end-to-end solution

Brief introduction to neural networks

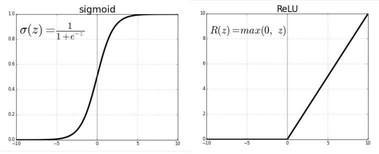
Deep Neural Networks (DNN)



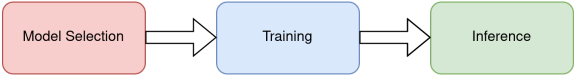
- Bengio, Hinton, LeCun (2009)
- big data + GPUs/TPUs
- learning with millions of neurons
- new architectures available for computer vision, video processing, NLP

DNN Techniques

- ReLU activation function
- dropout - type of regularization
- learning with mini-batches
- transfer learning



Neural Networks Life cycle



Model Selection

- find suitable architecture for the given problem
- neural architecture search (NAS)
- evolutionary algorithm, Bayesian optimisation, gradient based

Training

- find suitable weights for the given architecture and problem
- gradient approaches

Inference

- evaluating the final trained network

Model/Architecture Selection

Past

- how well the network performs on the given task
- accuracy on the test set
- the better accuracy the better network

Today

- enable inference on mobile devices
- multi-objective optimisation problem
- trade-off between accuracy and network complexity (size, energy and memory consumption, etc.)

AppNeCo project

- energy complexity of deep neural networks (Kalina, Šíma, Vidnerová)

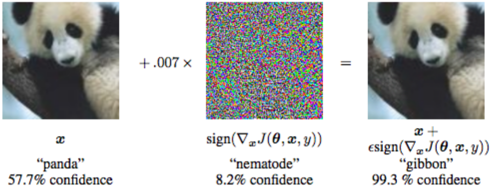
More objectives - robustness

Robustness objective

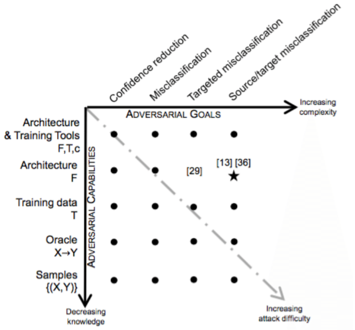
- need for robust models
- robustness to outliers, noise, adversarial examples

Adversarial examples

- perturbed examples (inputs) constructed to force the network to give a wrong answer



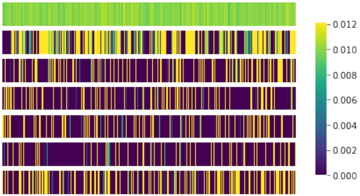
Adversarial attacks



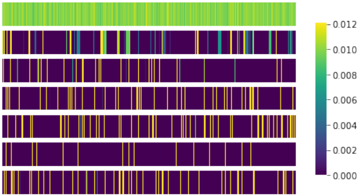
Robustness and variances

AlexNet

Pretrained network (left)

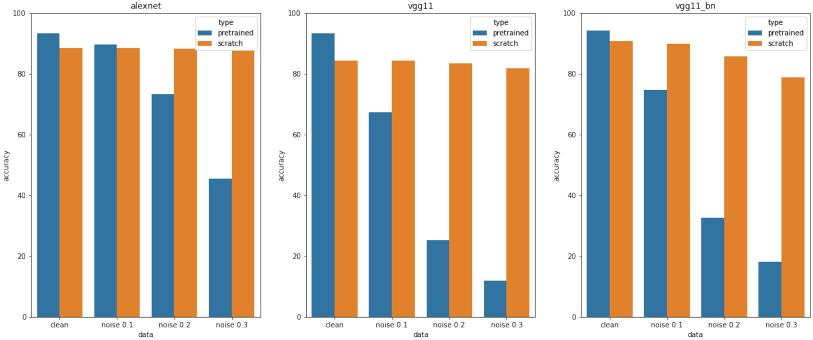


Trained from scratch (right)

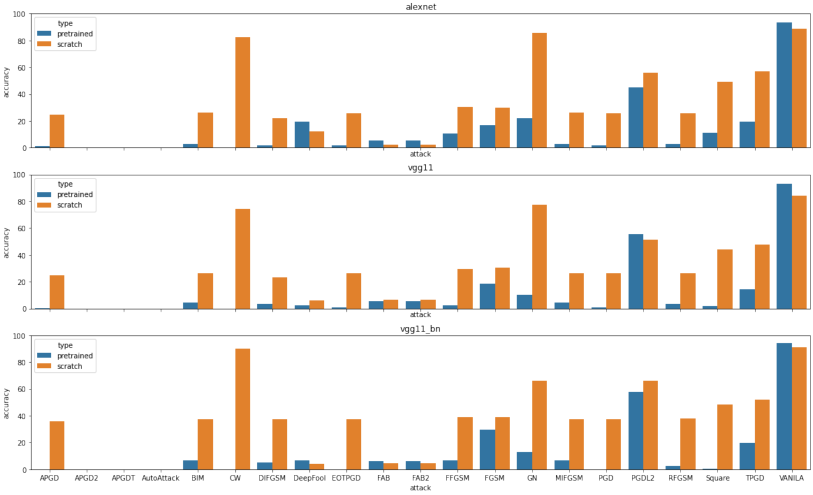


- brighter colour, higher variance

Robustness to noise



Robustness to adversarial examples

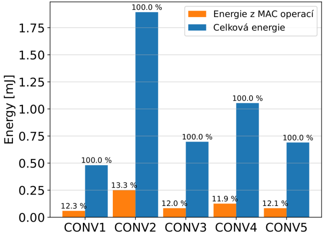


Energy efficient DNNs

Energy

$$E = E_{data} + E_{MAC}$$

(MAC ... multiply and accumulate)
E_{data} dominates



Methods

- reducing precision (quantisation), mixed precision
- pruning networks (sparse networks)
- approximate computations

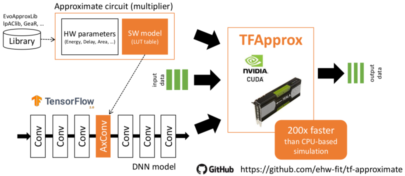
Approximate Computations

Approximate Adders and Multipliers

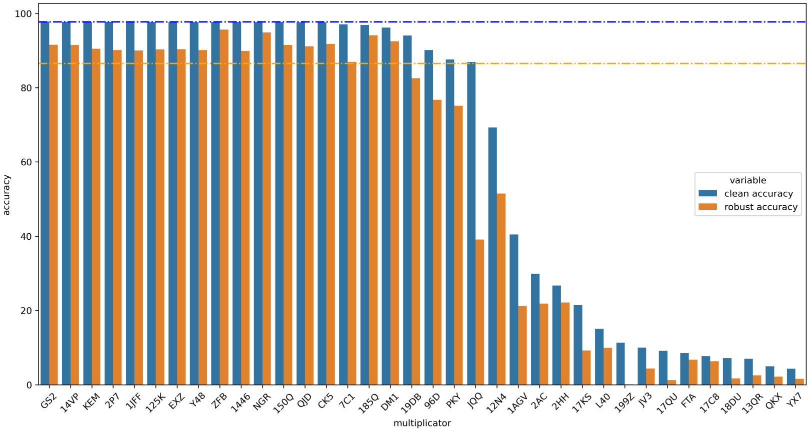
- work of our colleagues from Brno (Mrázek, Sekanina, Vašíček)
- evolutionary hardware - evolving approximate circuits
- multi-objective optimisation - error, energy, area and delay

Library tf-approximate

- library with Tensorflow interface
- provides approximate convolutional layers
- enables simulation of approximate computations



Approximate Layers - Robustness



Neural Architecture Search

Towards Multi-objectivity

- accuracy
- robustness
- model size
- energy

Reducing computational complexity

- NAS is typically very computationally demanding
- reducing time and energy consumption $\rightarrow$ "*green autoML*"
- performance prediction
- multi-objective performance prediction (Neruda, Kadlecová, Vidnerová, Pilát, Lukasik)

Reducing cost of NAS

Performance prediction task

- no need for exact prediction
- comparison between two models enough (one target)

Approaches

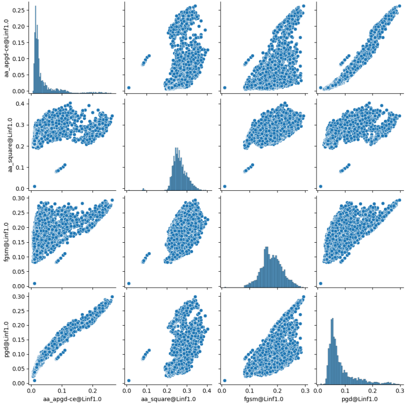
- surrogate models
- zero-cost proxies
- learning curve extrapolation

Our Goal

- performance prediction for diverse objectives
- multi-objective performance prediction

Predicting robustness

- evaluation of robustness is very time consuming
- prediction seems to be more difficult than prediction of accuracy



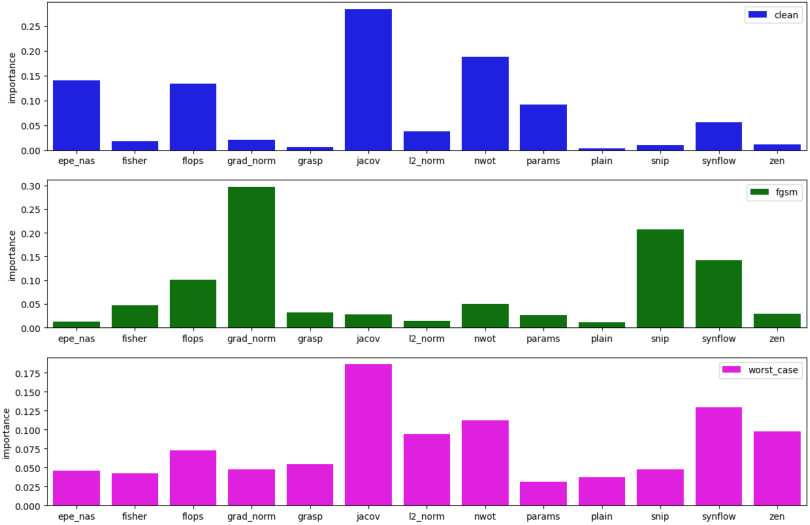
Prediction based on zero-cost proxies

- XGBoost, on NAS-Bench-201, 6466 networks
- cifar-100
- predict clean and robust accuracy based on zero-cost-proxies
- inspect feature importance
- 100/1000 training samples

training samples	avg r2 score					
	clean	**apgd**	**square**	**fgsm**	**pgd**	**worst case**
100	0.697	0.261	0.329	0.668	0.227	0.261
1000	0.892	0.513	0.579	0.813	0.509	0.514

Prediction based on zero-cost proxies

Feature importance



Conclusion

Takeaway

- finding a good model for the given task is a multi-objective problem
- main objectives:
  - accuracy
  - robustness
  - size
  - energy
- need for speed-up of the whole process

Thank you! Questions?